Supporting information：

Reacon: a template- and cluster-based framework for reaction condition prediction

Zihan Wang[‡a], *Kangjie Lin*[‡a], *Jianfeng Pei*[*b] *and Luhua Lai*[*ab]

a BNLMS, Peking-Tsinghua Center for Life Sciences, College of Chemistry and Molecular Engineering, Peking University, Beijing, 100871, China.

b Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, 100871, China.

* Corresponding authors, E-mail: lhlai@pku.edu.cn, jfpei@pku.edu.cn

‡ These authors contributed equally

Section 1. Clustering algorithm

Figure S1 illustrates the workflow of the condition clustering algorithm. For a reaction that needs to be classified, we first label it and then determine if it belongs to an existing cluster. If not, we establish a new cluster with the reaction label as the cluster label. If it belongs to an existing cluster, the reaction is added to the most appropriate existing cluster. After clustering, the labels of each cluster are updated using the following formula:

$$L_i = \{l \mid \frac{n(l)}{N} > \xi\} \quad (0 < \xi < 1) \tag{1}$$

The cluster label for each cluster comprises reaction condition labels that surpass a certain threshold of occurrence frequency. Here, L_i denotes the set of labels corresponding to the ith cluster, l represents a specific label, n(l) signifies the number of times this label appears in the reaction conditions within the cluster, N denotes the total number of reactions in the cluster, and ξ is the threshold value. In this study, we set ξ to 0.8 after observing the impact of different threshold on clustering effectiveness.

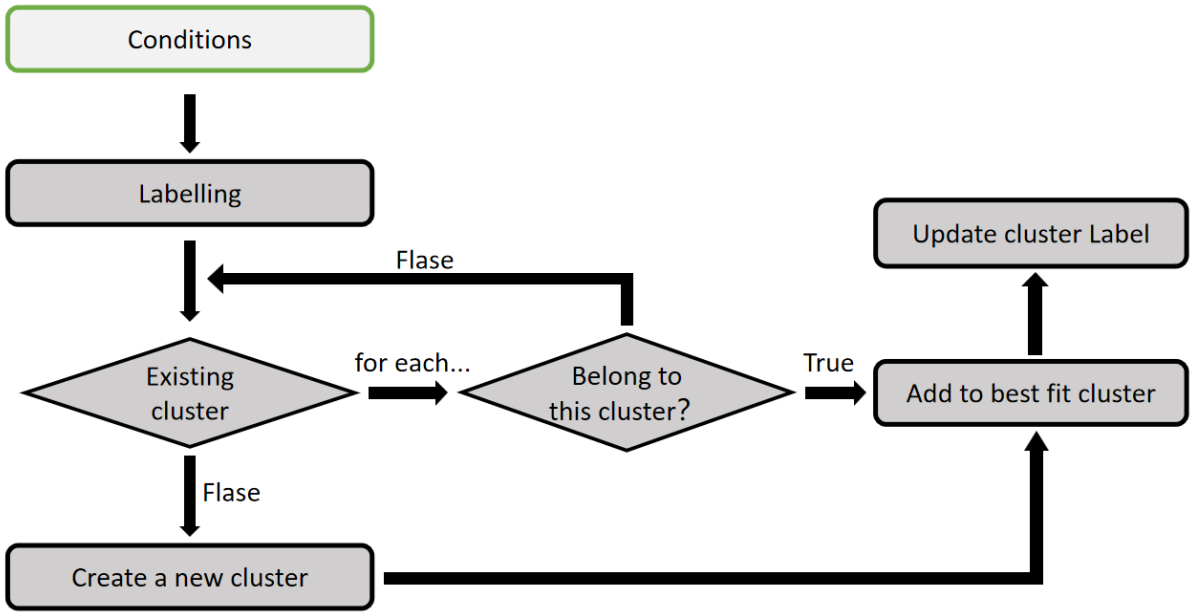

Figure S1. Workflow of the condition clustering algorithm

The labels used for extracting features and their corresponding pattern SMARTS could be

found in Table S1.

Table S1. Labels used for extracting features and their corresponding pattern SMARTS.

classification basis	Labels	Pattern Smarts
functional group	alkene	[CX3:1]=[CX3:2]
	alkyne	[CX2:1]#[CX2:2]
	alcohol	[CX4:1][OX2H]
	ether	[CX4:1][OX2][CX4:2]
	aldehyde	[CX3H1:1]=[OX1:2], [CX3H2:1]=[OX1:2]
	ketone	[#6][CX3H0:1](=[OX1])[#6]
	arboxylic acid	[!O:1][CX3:2](=[OX1])[OX2H1:3]
	ester	[CX3:1](=[OX1])[OX2H0:2]
	amide	[CX3:1](=[OX1])[NX3H2:2]
	nitro	[CX4:1][N+]([O-])=O
	amine/ammonia	[NX3:1], [n]
	halide	[F;H0;X1], [Cl;H0;X1], [Br;H0;X1], [I;H0;X1]
	acid chloride	[CX3:1](=[OX1])[Cl,Br,I:2]
	anhydride	[CX3:1](=[OX1])[OX2:2][CX3:3](=[OX1])
	nitrile	[NX1:1]#[CX2+0:2]
	aromatic	a
	Sulfone/sulfoxide	[SX4](=[OX1])(=[OX1]), [SX3](=[OX1])
	phosphine	[PX3], [PX4], [PX5], [PX6]
	metal alkyl	[CX4:1][Mg,Al,Zn,Li:2]
	silane	[SiX4:1][#6:2]
	sulfide	[CX4:1][SX2:2]
	alkane	[CX4:1]
element	transition metal	[#21,#22,#23,#24,#25,#26,#27,#28,#29,#30,#39,#40,#41,#42,#44,#45,#46,#47,#48,#72,#73,#74,#75,#76,#77,#78,#79,#80,#104,#105,#106,#107,#108,#109,#110,#111,#112]
	reducing metal	[#3,#4,#11,#12,#13,#19,#20,#30,#26;+0]
	Main group metal	[#13,#31,#49,#81,#50,#82,#83;]
function	oxidizer	[Cr+6], [Mn+7], [Mn+4], [Ce+4], [Pb+4], [I+3], O[N+]([O-])=O
	reductant	[H-], [BH4-], [AlH4-], [NaH], [LiH], [BH3], [BH2], [BH], [AlH3], [AlH2], [AlH]
	acid	[ClH1,BrH1,IH1:1], O=S(=O)(O)O, O[N+]([O-])=O, [CX3:1](=[OX1])[OX2H1:2]
	base	[OH1-], [NH2-], [NH1-], [NH0-], [SH-], [O-],[N-],[S-]
	lewis acid	[AlX3:1][F,Cl,Br,I,C:2],[BX3:1][F,Cl,Br,I,H:2],[Al+3], [Ti+4], [Zn+2], [ZnX2:1][Cl,Br,I:2], [Si+4], [Fe+3], [FeX3:1][Cl,Br,F:2], [Ge+4], [Sn+4], [Ce+4]
else	ionic	[+,-]

Table S2 shows attribution of reaction conditions under the templates corresponding to the reactions in Figure 4A. Each cluster includes its corresponding cluster label and the reaction conditions it contains. Cluster labels are divided into two parts: the first part represents the catalyst label, and the second part represents the solvent and reagent labels. Additionally, it is noteworthy that the conditions in clusters 9, 10, and 11 are clearly erroneous. These errors mainly stem from data loss while extractions, which were initially hidden within a large amount of condition data and difficult to detect. However, through clustering, these errors are identified and separated accordingly.

Table S2. The attribution of reaction conditions under one template after clustering.

Cluster id	Cluster labels	Conditions
Cluster 1	['transition metal', 'halide'], ['silane', 'halide']	'Cl[Ti](Cl)(Cl)Cl', 'ClCCl', 'None', 'CC[SiH](CC)CC', 'None', 'None'
Cluster 2	[], ['reductant', 'ether', 'lewis acid']	'None', 'C1CCOC1', 'None', '[Al+3].[Cl-].[Cl-].[Cl-]', '[BH4-].[Na+]', 'None'
Cluster 3	[],[‘amine/ammonia’, ’alcohol’,’base’]	'None', 'CCO', 'None', 'NN.O', 'None', 'None'
		'None', 'CCOCC', 'O', '[K+].[OH-]', 'NN.O', 'OCCOCCOCCO'
		'None', 'O', 'None', '[K+].[OH-]', 'NN.O', 'OCCOCCO'
		'None', 'O', 'OCCOCCO', '[Na+].[OH-]', 'NN.O', 'None'
		'None', 'OCCO', 'None', '[K+].[OH-]', 'NN.O', 'None'
		'None', 'OCCOCCO', 'None', 'NN.O', 'None', 'None'
		'None', 'OCCOCCO', 'None', '[K+].[OH-]', 'Cl', 'NN.O'
		'None', 'OCCOCCO', 'None', '[K+].[OH-]', 'NN', 'None'
		'None', 'OCCOCCO', 'None', '[K+].[OH-]', 'NN.O', 'None'
		'None', 'OCCOCCO', 'None', '[Na+].[OH-]', 'NN', 'None'
		'None', 'OCCOCCO', 'None', '[Na+].[OH-]', 'NN.O', 'None'
		'None', 'OCCOCCOCCO', 'None', '[K+].[OH-]', 'NN.O', 'None'
		'None', 'CCO', 'None', 'NN.O', 'None', 'None'
Cluster 4	[], ['halide', 'carboxylic acid', 'acid', 'silane']	'None', 'CCCCCC', 'None', 'CC[SiH](CC)CC', 'O=C(O)C(F)(F)F', 'None'
		'None', 'CC[SiH](CC)CC', 'O=C(O)C(F)(F)F', 'None', 'None', 'None'
		'None', 'None', 'None', 'CC[SiH](CC)CC', 'O=C(O)C(F)(F)F', 'None'
		'None', 'None', 'None', 'O=C([O-])O.[Na+]',

		'CC[SiH](CC)CC', 'O=C(O)C(F)(F)F'
		'None', 'None', 'None', '[Na+].[OH-]', 'CC[SiH](CC)CC', 'O=C(O)C(F)(F)F'
		'None', 'O', 'None', 'CC[SiH](CC)CC', 'O=C(O)C(F)(F)F', 'None'
		'None', 'CC(C)O', 'None', 'NC1CCCCC1', 'None', 'None'
		'None', 'O=C(O)C(F)(F)F', 'None', 'CC[SiH](CC)CC', 'None', 'None'
		'None', 'O=C(O)C(F)(F)F', 'None', 'CC[SiH](CC)CC', 'O', 'None'
		'None', 'O=C(O)C(F)(F)F', 'None', 'None', 'None', 'None'
		'None', 'O=C(O)C(F)(F)F', 'None', 'O=C([O-])O.[Na+]', 'CC[SiH](CC)CC', 'None'
		'None', 'O=C(O)C(F)(F)F', 'None', '[Na+].[OH-]', 'CC[SiH](CC)CC', 'None'
Cluster 5	['transition metal'], ['alcohol']	'[C].[Pd]', 'CO', 'None', '[H][H]', 'None', 'None'
		'[Ni]', 'CO', 'None', 'None', 'None', 'None'
		'[Pd]', 'CCO', 'None', 'None', 'None', 'None'
		'[Pd]', 'CO', 'None', 'None', 'None', 'None'
		'[Pd]', 'CO', 'None', '[H][H]', 'None', 'None'
Cluster 6	['transition metal'], ['carboxylic acid', 'acid']	'[H][H].[Pd]', 'CC(=O)O', 'None', 'None', 'None', 'None'
		'[Pd]', 'CC(=O)O', 'None', 'None', 'None', 'None'
		'[Pd]', 'CC(=O)O', 'None', 'O=S(=O)(O)O', 'None', 'None'
		'[Pd]', 'CC(=O)O', 'None', '[H][H]', 'None', 'None'
		'[Pd]', 'CC(=O)O', 'None', '[O-][Cl+3]([O-])([O-])O', 'None', 'None'
		'[Pd]', 'CC(=O)O', 'None', '[O-][Cl+3]([O-])([O-])O', '[H][H]', 'None'
		'[Zn]', 'CC(=O)O', 'None', 'None', 'None', 'None'
Cluster 7	['transition metal'], ['alcohol', 'acid']	['[Pd]', 'CCO', 'None', 'Cl', 'None', 'None']
Cluster 8	['reducing metal', 'transition metal'], ['acid', 'aromatic']	'[Zn]', 'C1COCCO1', 'None', 'Cc1ccccc1', 'Cl', 'O'
		'[Zn]', 'Cc1ccccc1', 'None', 'Cl', 'O', 'None'
		'[Zn]', 'Cl', 'O', 'Cc1ccccc1', 'None', 'None'
		'[Zn]', 'O', 'None', 'Cc1ccccc1', 'Cl', 'None'
		'[Zn]', 'CC(=O)O', 'None', 'Cl', 'None', 'None'
Cluster 9	[], ['acid', 'aromatic']	'None', 'Cc1ccccc1', 'None', 'Cl', 'O', 'None'
Cluster 10	[], ['ether', 'acid']	'None', 'Cl', 'None', 'None', 'None', 'None'
		'None', 'O', 'None', 'Cl', 'None', 'None'
Cluster 11	[], ['ether', 'acid']	'None', 'O', 'None', 'C1COCCO1', 'Cl', 'None'

Section 2. Model Details.

D-MPNN Model. D-MPNN[1] is a variant of the Message Passing Neural Network (MPNN), which is essentially a type of Graph Convolutional Neural Network (GCNN). Unlike traditional MPNN, which associate messages with vertices, D-MPNN associates messages with directed edges rather than vertices. The initial directed edge feature h_{vw}^{0} is constructed by concatenating the feature x_v of the first atom of the bond to the respective bond feature e_{vw} and then passing the concatenated vector through a single neural network layer.

$$h_{\mathrm{vw}}^{0} = \tau\left(\mathbf{W}_i \,\mathrm{cat}(x_v, e_{vw})\right) \tag{2}$$

Here $\mathbf{W}_i \in \mathbb{R}^{h \times h_i}$ is a learned matrix, h represents the size of the hidden layer, and h_i is the size of the concatenated vector $\mathrm{cat}(x_v, e_{vw})$. τ is a nonlinear activation function. Then, the directed edge features are updated through T steps of message passing.

$$h_{vw}^{\mathrm{t+1}} = \tau\left(h_{vw}^{0} + \mathbf{W}_h \sum_{\mathrm{k}\in\{\mathrm{N}(v)\setminus w\}} h_{kv}^{\mathrm{t}}\right) \tag{3}$$

Here $\mathbf{W}_h \in \mathbb{R}^{h \times h}$ is a learned matrix, and $\mathrm{N}(v)\setminus w$ denotes the set of neighbors of node v excluding *w*.

GAT Model. The Graph Attention Network (GAT)[2] is a type of neural network architecture that is specifically designed to operate on graph-structured data. It fundamentally enhances node feature aggregation by computing attention coefficients that determine the significance of each neighbor's contribution. The raw attention scores between node *i* and node *j* are computed using:

$$e_{ij} = \mathrm{LeakyReLU}\left(\alpha^{\mathrm{T}}\mathrm{cat}\left(\mathbf{W}h_i, \mathbf{W}h_j\right)\right) \tag{4}$$

Here α is a learnable weight vector, W is a weight matrix applied to every node's features. Then the softmax function normalizes the attention scores e_{ij} across all neighbors *k* of node *i* resulting in normalized attention coefficients α_{ij} that sum to one.

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathrm{N}_i} \exp(e_{ik})} \tag{5}$$

These coefficients are then used to weight the neighbors' feature contributions during aggregation.

The inputs to the model utilize the same node features and bond features as those employed in the D-MPNN model.

Section 3. Model training and prediction results

Algorithm 1: Reaction condition prediction workflow

input : Input reaction **r**; Corresponding template **t**.
output: A sorted dictionary containing the predicted conditions and their Conscore.

```
ClusteredLib ← Cluster(Lib);
// Lib is the Template-Reaction library, input in
   dictionary form.This operation clusters the reactions
   under each template
for model, component in M do
    pred[component] ← Prediciton(model, r);
    // Get the predictions for each individual model.M is a
       set of models used to predict each component of the
       reaction conditions.
end
candidates ← GetCandidates(Lib, t);
// Query the template-condition library to obtain the
   reaction conditions recorded under the corresponding
   template.
if candidates is None then
    for component in pred do
        Top3[component] = GetTopK(pred[component], 3)
    end
    candidates ← GenerateCandidates(Top3);
    // Here we will randomly combine the Top-3 of different
       condition components to generate the final candidate
end
for condition in candidates do
    ConScore ← CalConScore(pred, condition);
    ScoreDict[condition] ← ConScore;
    // Calculate the Conscore for each candidate condition
end
ClusterScoreDict ← CalClusterScore(ScoreDict, ClusteredLib);
// The value of ClusterScore is equal to the highest
   ConScore in each cluster
return SortByValue(ScoreDict), SortByValue(ClusterScoreDict)
```

Figure S2. Pseudo-code for template-based reaction prediction.

The hyperparameters of the final model are determined experimentally. We constructed a subset containing 10% of the dataset of the training set and trained models on it using different

hyperparameters. Eventually we tested the performance of the models on the validation set. We finally selected the best performing set of hyperparameters for final use. The range of hyperparameters tested is shown below. The optimal parameters are in bold.

Table S3. The hyperparameters optimization of GAT model.

Parameters	Space
Depth	{3, 4, **5**}
Dropout rate	{0, 0.1, **0.2**, 0.3}
Heads	{1, 4, **8**, 16}
Learning rate	{0.01, **0.001**, 0.0001}
FFN layers	{**1**, 2}

Table S4. The hyperparameters optimization of D-MPNN model.

Parameters	Space
Depth	{**3**, 4, 5}
Dropout rate	{**0**, 0.1, 0.2, 0.3}
hidden size	{100, 200, **300**}
Learning rate	{0.01, **0.001**, 0.0001}
FFN layers	{1, **2**}

Table S5. The individual prediction performance of different models on each component of reaction conditions in the test set.

Models	Catalyst	Solvent 1	Solvent 2	Reagent 1	Reagent 2	Reagent 3
	Top-3 (%)	Top-3 (%)	Top-3 (%)	Top-3(%)	Top-3 (%)	Top-3 (%)
Popularity baseline	96.91	80.01	89.78	85.24	84.37	95.44
MLP	90.23	46.27	89.14	46.42	80.41	95.60
Similarity baseline	92.47	57.26	90.23	70.12	82.33	96.01
D-MPNN	**98.62**	**85.44**	**95.56**	**89.35**	**91.72**	**98.51**
D-MPNN (multi-task)	98.28	83.79	94.78	88.23	90.93	98.34

Section 4. Additional model performance

Table S6. The individual prediction performance of different models on each component of reaction conditions in the time-split test set.

Models	Catalyst	Solvent 1	Solvent 2	Reagent 1	Reagent 2	Reagent 3
	Top-1 (%)	Top-1 (%)	Top-1 (%)	Top-1 (%)	Top-1 (%)	Top-1 (%)
Popularity baseline	86.04	34.05	80.84	37.98	70.21	94.29
MLP	85.77	33.19	80.54	36.35	69.50	94.08
Similarity baseline	85.92	29.14	79.16	34.70	69.47	94.21
D-MPNN	**87.35**	**57.23**	**83.59**	**58.08**	**74.71**	**94.68**

Table S7. Performance of different models on condition predictions in the time-split test set.

Metric	Model	Top-1 (%)	Top-3 (%)	Top-10 (%)
Exact accuracy	Popularity baseline	22.10	37.36	51.92
	MLP	22.05	34.88	50.54
	Similarity baseline	18.83	33.51	48.90
	D-MPNN	**33.96**	**48.07**	**60.08**

Table S8. Performance of different models on condition predictions in the time-split test set after clustering.

Metric	Model	Top-1 (%)	Top-3 (%)	Top-10 (%)
Cluster accuracy	Popularity baseline	42.37	67.74	82.19
	MLP	42.17	66.84	81.70
	Similarity baseline	39.31	65.29	79.88
	D-MPNN	**57.43**	**75.61**	**84.73**

Table S9. The individual prediction performance of different models on each component of reaction conditions in the USPTO-Condition dataset.

Model	Catalyst	Solvent 1	Solvent 2	Reagent 1	Reagent 2
	Top-1 (%)	Top-1 (%)	Top-1 (%)	Top-1 (%)	Top-1 (%)
RCR	92.91	50.15	81.30	49.72	76.22
Parrot-LM-E	**92.50**	50.18	80.96	**50.39**	76.48
D-MPNN	92.44	**50.39**	**81.58**	50.02	**77.95**

Table S10. Performance of different models on condition predictions in the USPTO-Condition dataset.

Metric	Model	Top-1 (%)	Top-3 (%)	Top-10 (%)
Exact accuracy	RCR	25.96	37.71	46.12
	Parrot-LM-E	26.91	40.25	49.14
	D-MPNN	**27.52**	**43.90**	**58.63**

Section 5 Clustering results statistics

Table S11. displays the cluster sizes in our template-condition library, revealing that the majority of clusters are comprised of fewer than 10 conditions.

Table S11. Cluster size statistics.

Library	Cluster size			
	<3	3-10	>10	Mean
r1	45019	10347	1149	2.466
r0	19506	5948	1063	3.342
r0*	14940	5030	1014	3.692

There is an increasing trend in the size of the clusters from r1 to r0*, further we tested the clustering accuracy when using different libraries and the results are shown in Table S12.

Table S12. Cluster accuracy when using different library.

Library	Top-1 (%)	Top-3 (%)	Top-10 (%)
r0*	62.96	83.51	93.97
r0	63.27	83.57	94.18
r1	65.56	84.52	95.28

The size of the clusters is also largely affected by the clustering criteria we set. We have tried here the size of the clusters as well as their predictive performance when using stricter clustering criteria (requiring at least three more identical reagent or solvent labels) and looser clustering criteria (requiring only one identical reagent or solvent label)

Table S13. Cluster size and cluster accuracy when using different criteria.

Criteria	Mean r1 cluster size	Top-1 (%)	Top-3 (%)	Top-10 (%)
strict	2.107	55.55	75.83	90.51
original	2.466	65.68	85.65	96.11
loose	3.084	80.65	94.74	99.03

We notice that the cluster accuracy of the model shows a clear correlation with the size of the clusters. This is easy to understand, larger clusters make it easier to judge the model's predictions as the same type of condition with ground truth. However, we need to point out that too loose clustering criteria can cause the results of clustering to deviate from chemical intuition.

Section 6. Supplemental cases of incorrectly predicted conditions

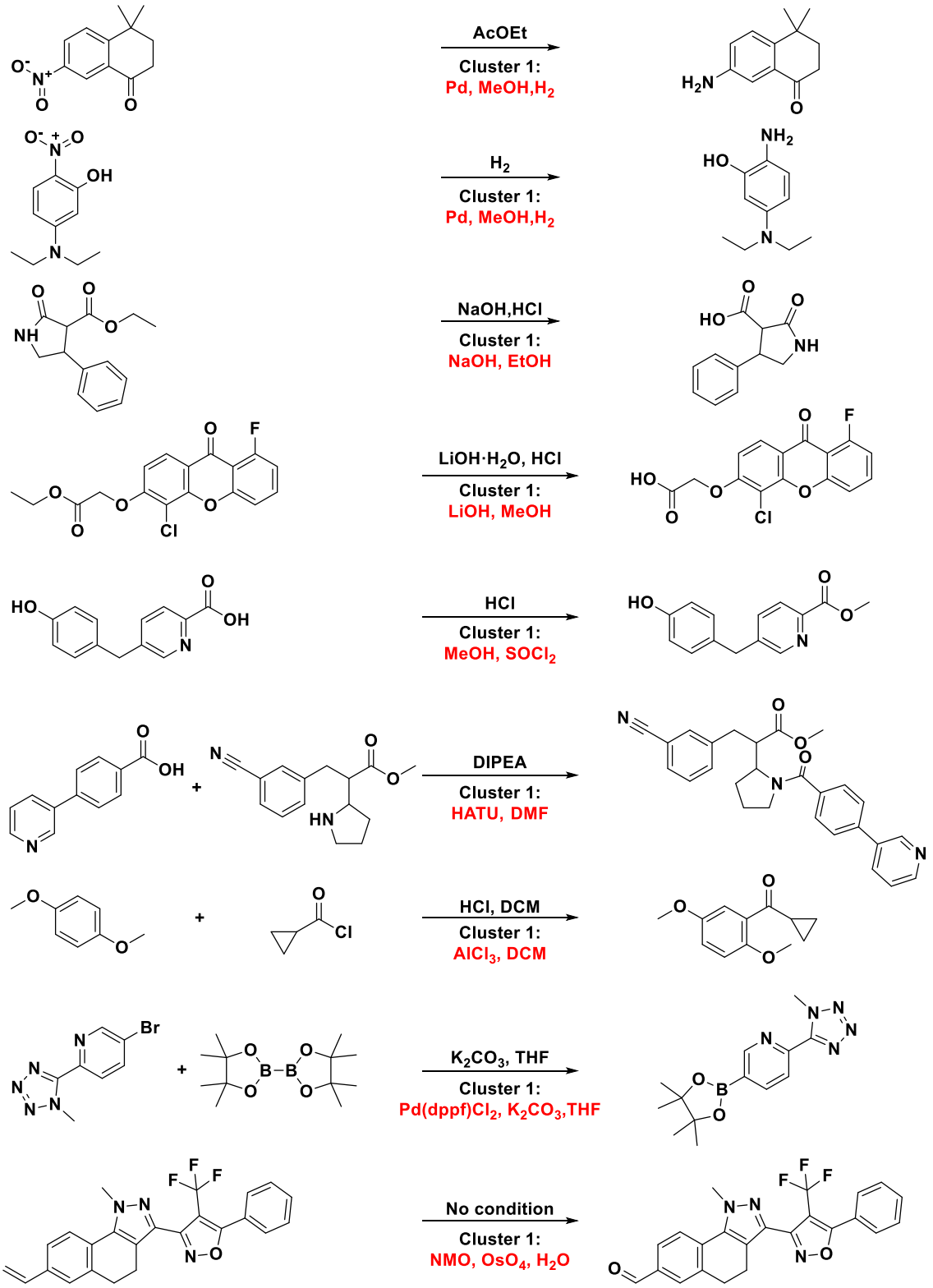

Figure S3. Reaction cases where our model's predictions do not agree with the ground truth. The conditions recorded in the dataset or in the literature are shown in black font in the figure, and the conditions predicted by the model are shown in red.

Section 7. Prediction of reaction conditions in actual drug synthesis routes

Regarding the reactions of recently synthesized drug molecules, different template libraries exhibit varying coverage ranges. For the 12 pathways comprising a total of 100 reactions selected from *Journal of Medicinal Chemistry*, the coverage rates of the r1, r0, and r0* template libraries are 0.54, 0.82 and 0.97, respectively.

Table S14. Overall predictive performance of our model on real routes. Here it is recorded in terms of the number of correctly predicted conditions/total conditions.

Molecular name	Exact accuracy		Cluster accuracy		
	Top-1	Top-3	Top-1	Top-3	Top-10
KB-0742[3]	0/4	0/4	2/4	2/4	3/4
LNP1892[4]	1/9	3/9	7/9	8/9	9/9
GDC-1971[5]	3/13	3/13	9/13	10/13	11/13
IPG7236[6]	3/7	5/7	4/7	6/7	7/7
GSK2982772[7]	1/6	3/6	6/6	6/6	6/6
MK-8189[8]	3/5	3/5	4/5	5/5	5/5
RP-6306[9]	2/5	2/5	2/5	3/5	4/5
AZD4831[10]	3/9	5/9	4/9	9/9	9/9
a1[11]	2/8	3/8	6/8	7/8	7/8
HPG1860 [12]	2/6	3/6	5/6	6/6	6/6
SAGE-718[13]	2/9	4/9	6/9	8/9	8/9
PF-07258669[14]	3/19	5/19	14/19	16/19	17/19
Total	25/100	39/100	64/100	85/100	92/100

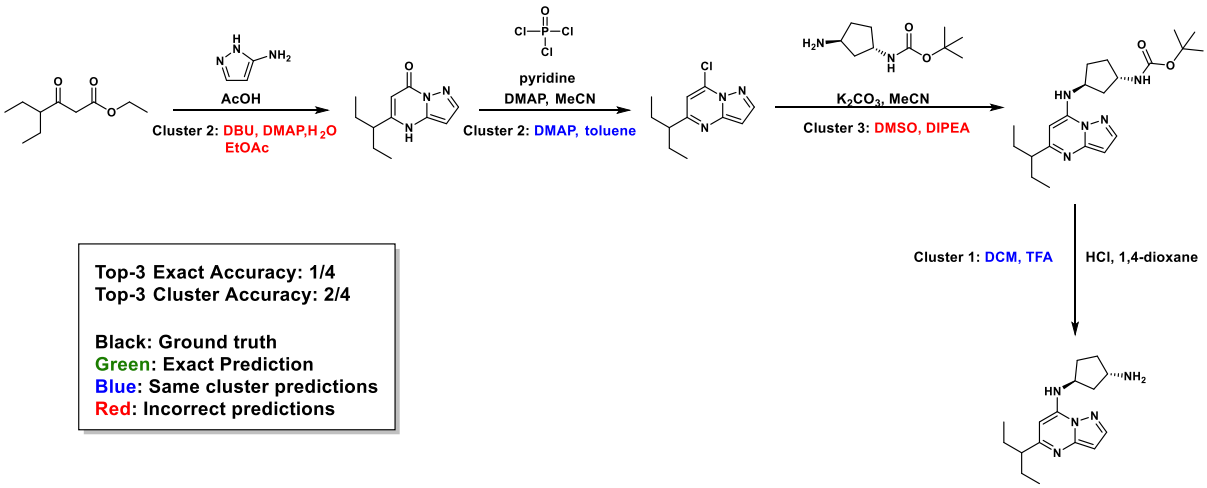

Figure S4. Synthesis route of KB-0742[3] with actual and predicted conditions.

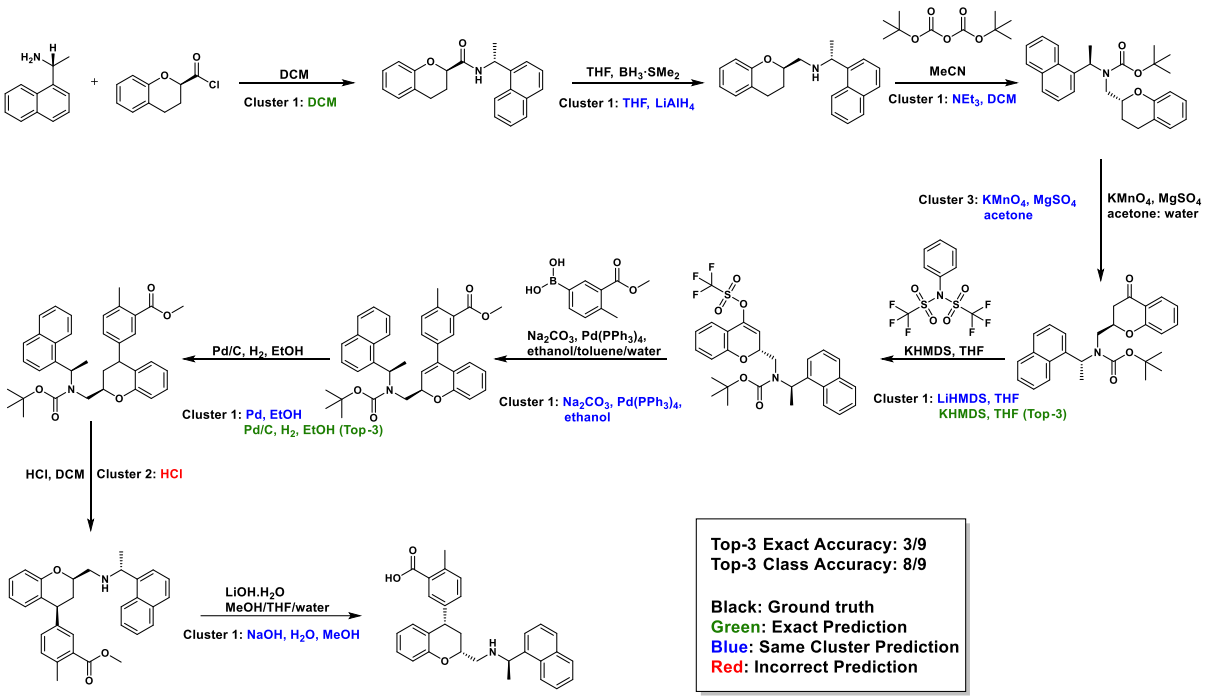

Figure S5. Synthesis route of LNP1892[4] with actual and predicted conditions.

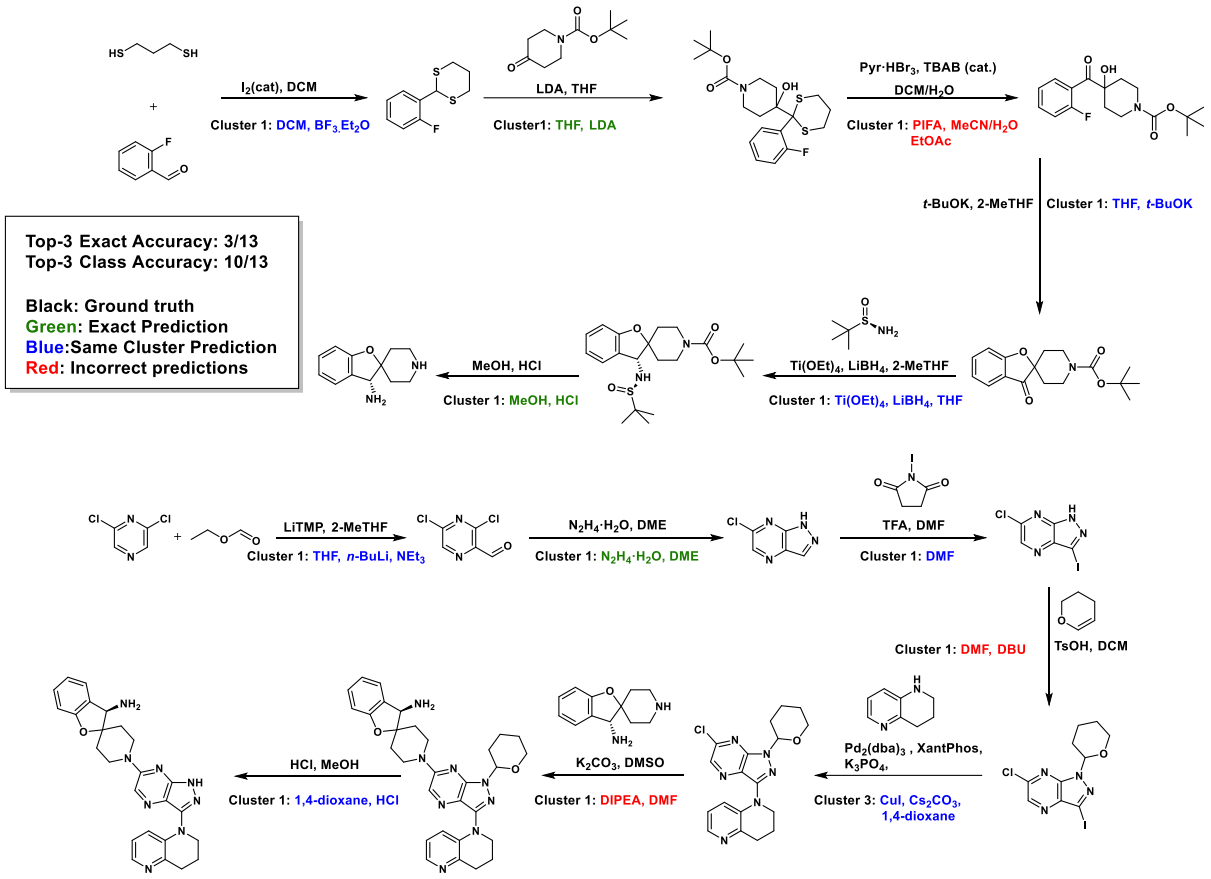

Figure S6. Synthesis route of GDC-1971[5] with actual and predicted conditions.

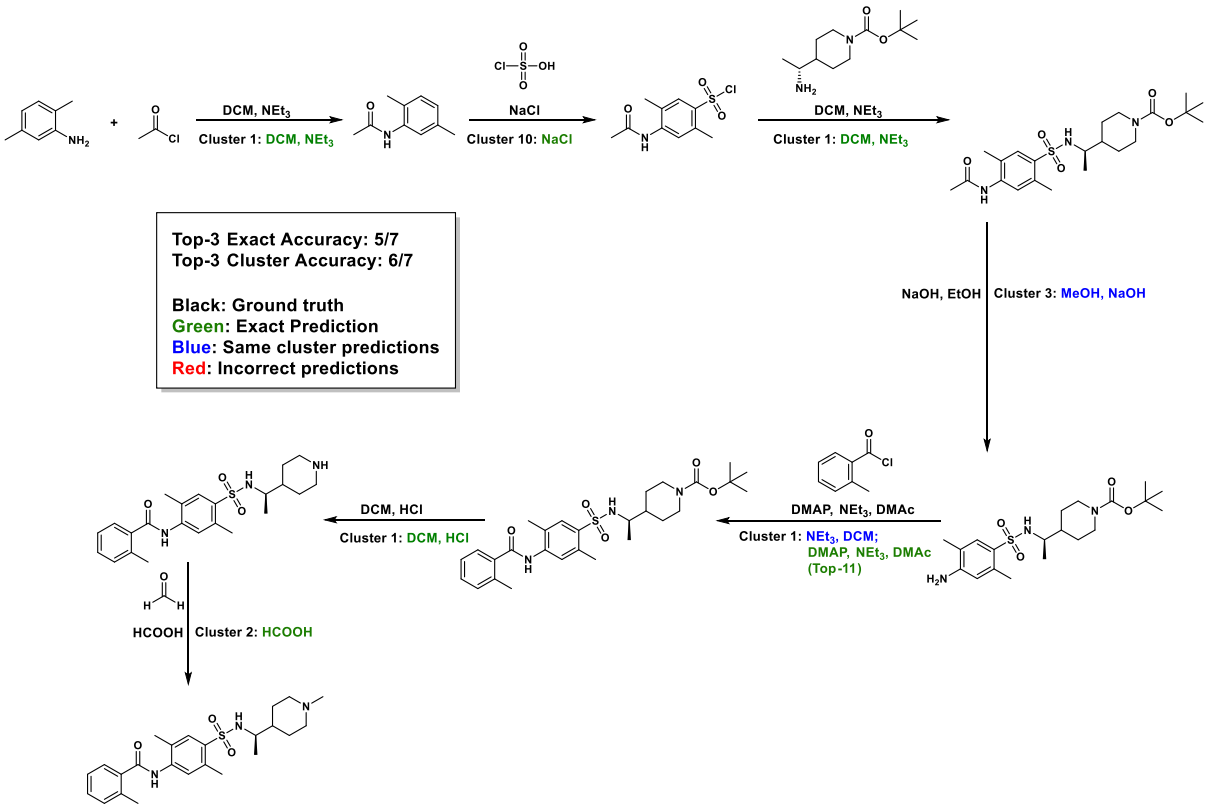

Figure S7. Synthesis route of IPG723[6] with actual and predicted conditions.

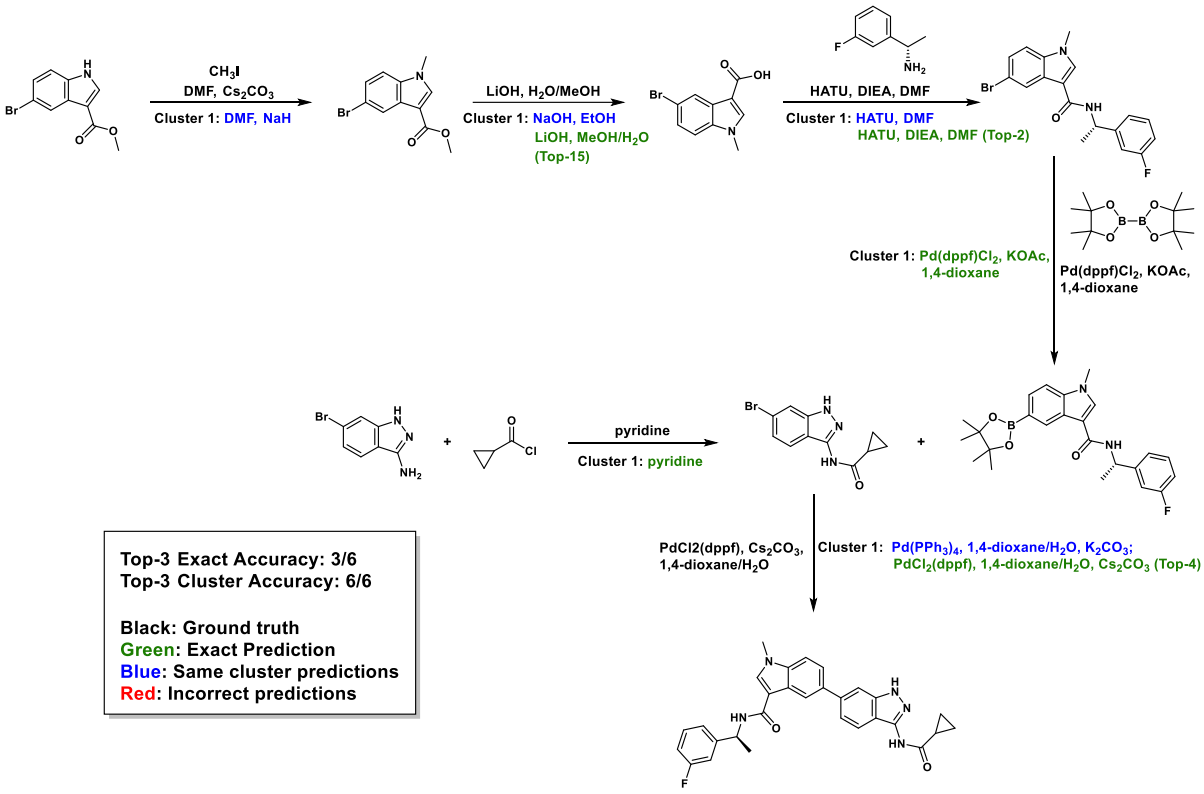

Figure S8. Synthesis route of GSK2982772 [7] with actual and predicted conditions.

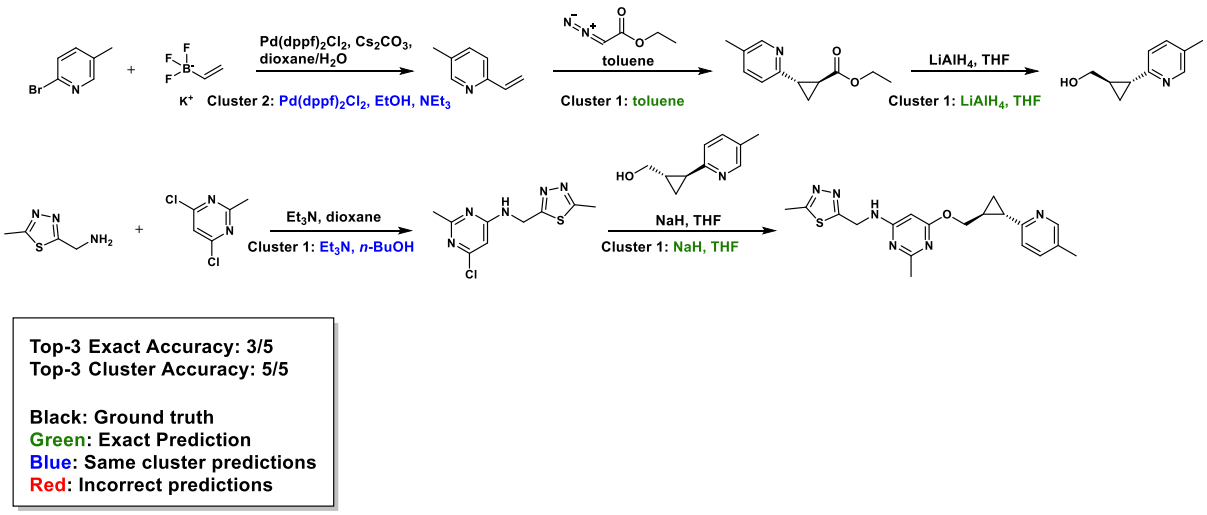

Figure S9. Synthesis route of MK-8189 [8] with actual and predicted conditions.

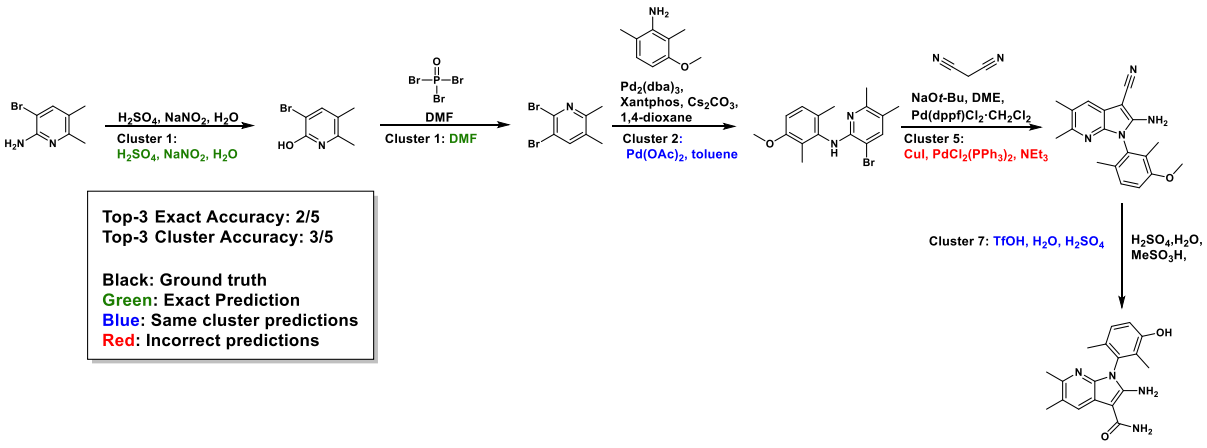

Figure S10. Synthesis route of RP-6306 [9] with actual and predicted conditions.

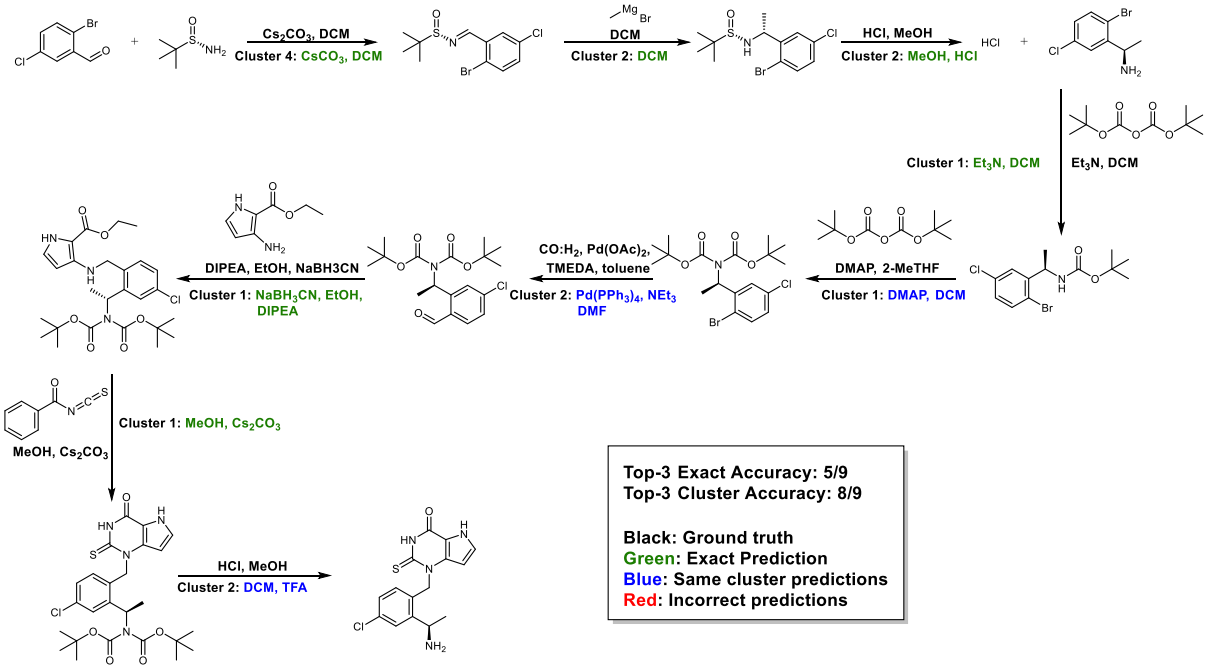

Figure S11. Synthesis route of AZD4831 [10] with actual and predicted conditions.

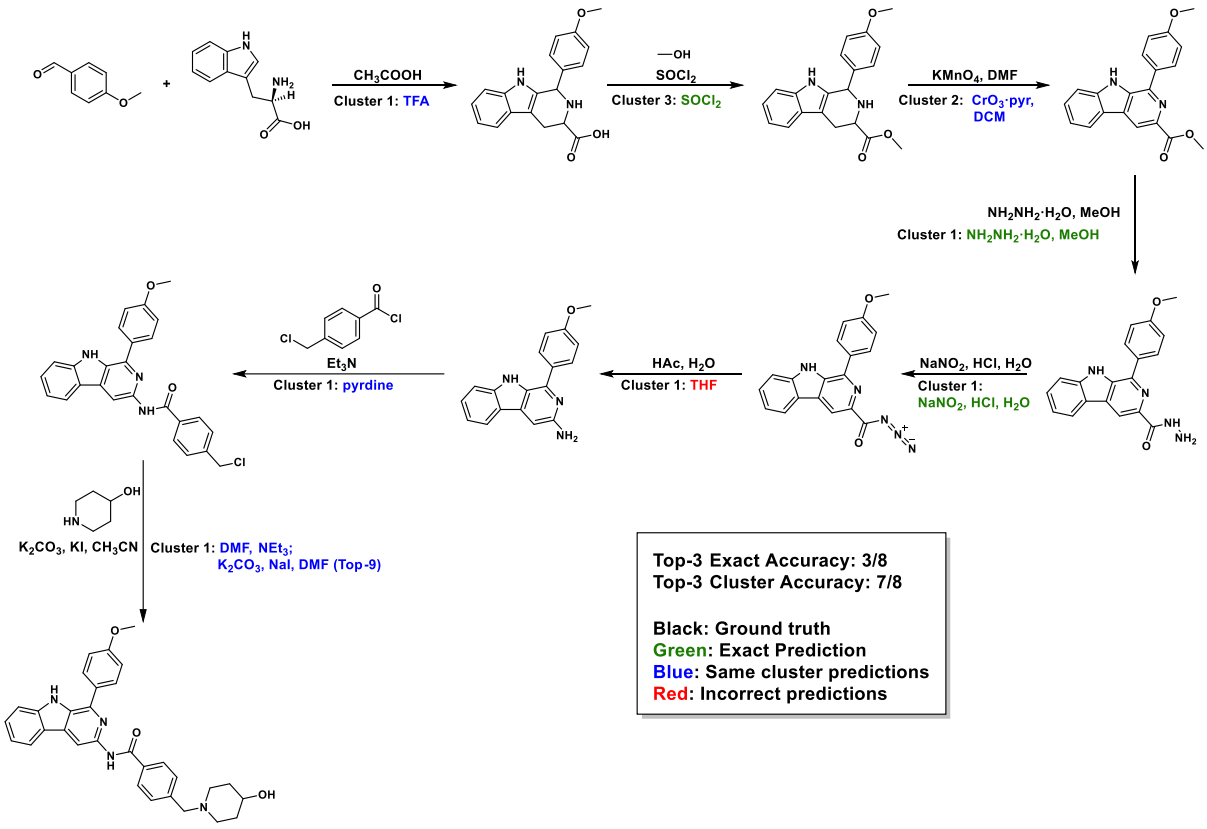

Figure S12. Synthesis route of 1,3-Substituted beta-Carboline [11] with actual and predicted conditions.

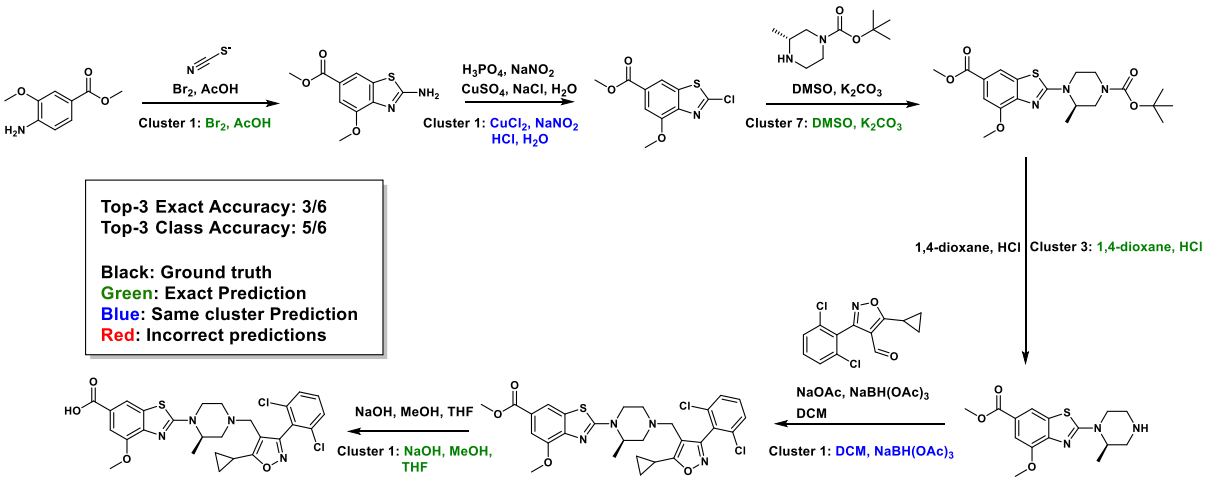

Figure S13. Synthesis route of HPG1860 [12] with actual and predicted conditions.

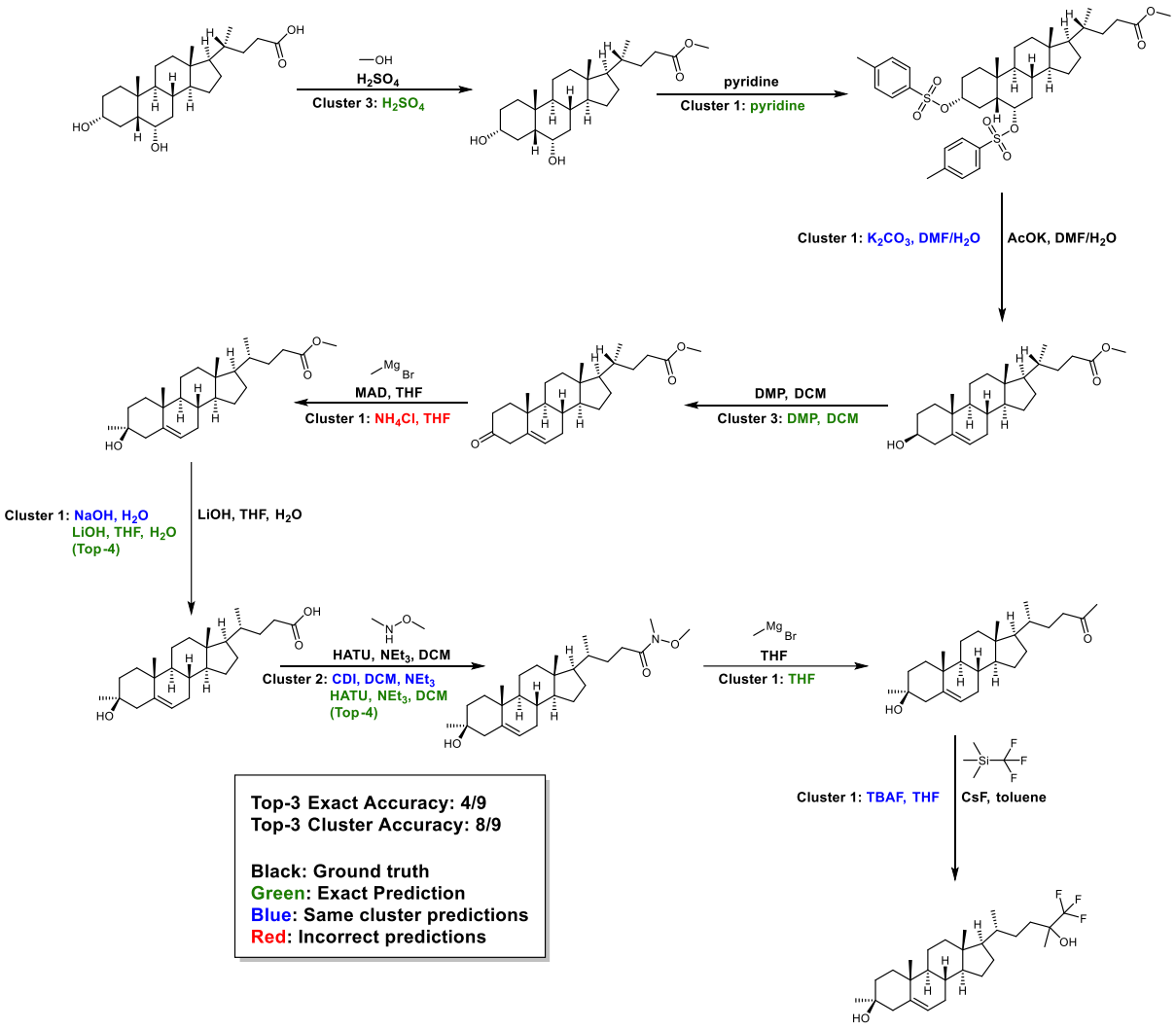

Figure S14. Synthesis route of SAGE-718 [13] with actual and predicted conditions.

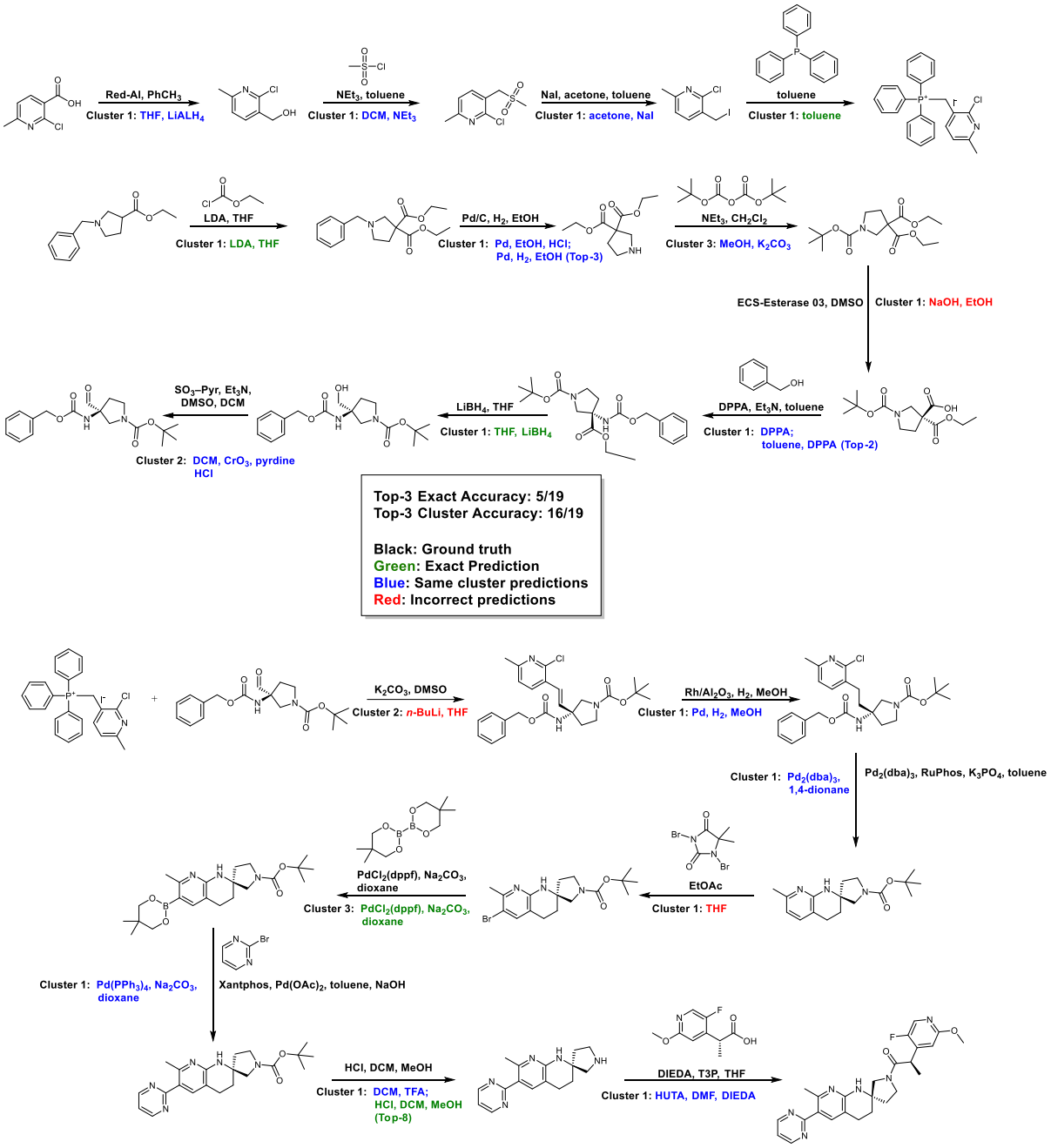

Figure S15. Synthesis route of PF-07258669 [14] with actual and predicted conditions.

REFERENCES

1. Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M.; Palmer, A.; Settels, V.; Jaakkola, T.; Jensen, K.; Barzilay, R., Analyzing Learned Molecular Representations for Property Prediction. *Journal of Chemical Information and Modeling* **2019,** *59* (8), 3370-3388.

2. Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio', P.; Bengio, Y., Graph Attention Networks. *ArXiv* **2017,** *abs/1710.10903*.

3. Freeman, D. B.; Hopkins, T. D.; Mikochik, P. J.; Vacca, J. P.; Gao, H.; Naylor-Olsen, A.; Rudra, S.; Li, H.; Pop, M. S.; Villagomez, R. A.; Lee, C.; Li, H.; Zhou, M.; Saffran, D. C.; Rioux, N.; Hood, T. R.; Day, M. A. L.; McKeown, M. R.; Lin, C. Y.; Bischofberger, N.; Trotter, B. W., Discovery of KB-0742, a Potent, Selective, Orally Bioavailable Small Molecule Inhibitor of CDK9 for MYC-Dependent Cancers. *J Med Chem* **2023,** *66* (23), 15629-15647.

4. Shukla, M. R.; Sadasivam, G.; Sarde, A.; Sayyed, M.; Pachpute, V.; Phadtare, R.; Walke, N.; Chaudhari, V. D.; Loriya, R.; Khan, T.; Gote, G.; Pawar, C.; Tryambake, M.; Mahajan, N.; Gandhe, A.; Sabde, S.; Pawar, S.; Patil, V.; Modi, D.; Mehta, M.; Nigade, P.; Modak, V.; Ghodke, R.; Narasimham, L.; Bhonde, M.; Gundu, J.; Goel, R.; Shah, C.; Kulkarni, S.; Sharma, S.; Bakhle, D.; Kamboj, R. K.; Palle, V. P., Discovery of LNP1892: A Precision Calcimimetic for the Treatment of Secondary Hyperparathyroidism. *J Med Chem* **2023,** *66* (14), 9418-9444.

5. Taylor, A. M.; Williams, B. R.; Giordanetto, F.; Kelley, E. H.; Lescarbeau, A.; Shortsleeves, K.; Tang, Y.; Walters, W. P.; Arrazate, A.; Bowman, C.; Brophy, E.; Chan, E. W.; Deshmukh, G.; Greisman, J. B.; Hunsaker, T. L.; Kipp, D. R.; Saenz Lopez-Larrocha, P.; Maddalo, D.; Martin, I. J.; Maragakis, P.; Merchant, M.; Murcko, M.; Nisonoff, H.; Nguyen, V.; Nguyen, V.; Orozco, O.; Owen, C.; Pierce, L.; Schmidt, M.; Shaw, D. E.; Smith, S.; Therrien, E.; Tran, J. C.; Watters, J.; Waters, N. J.; Wilbur, J.; Willmore, L., Identification of GDC-1971 (RLY-1971), a SHP2 Inhibitor Designed for the Treatment of Solid Tumors. *J Med Chem* **2023,** *66* (19), 13384-13399.

6. Wu, Y.; Xi, J.; Li, Y.; Li, Z.; Zhang, Y.; Wang, J.; Fan, G. H., Discovery of a Potent and Selective CCR8 Small Molecular Antagonist IPG7236 for the Treatment of Cancer. *J Med Chem* **2023,** *66* (7), 4548-4564.

7. Zhang, L.; Li, Y.; Tian, C.; Yang, R.; Wang, Y.; Xu, H.; Zhu, Q.; Chen, S.; Li, L.; Yang, S., From Hit to Lead: Structure-Based Optimization of Novel Selective Inhibitors of Receptor-Interacting Protein Kinase 1 (RIPK1) for the Treatment of Inflammatory Diseases. *J Med Chem* **2024,** *67* (1), 754-773.

8. Layton, M. E.; Kern, J. C.; Hartingh, T. J.; Shipe, W. D.; Raheem, I.; Kandebo, M.; Hayes, R. P.; Huszar, S.; Eddins, D.; Ma, B.; Fuerst, J.; Wollenberg, G. K.; Li, J.; Fritzen, J.; McGaughey, G. B.; Uslaner, J. M.; Smith, S. M.; Coleman, P. J.; Cox, C. D., Discovery of MK-8189, a Highly Potent and Selective PDE10A Inhibitor for the Treatment of Schizophrenia. *J Med Chem* **2023,** *66* (2), 1157-1171.

9. Szychowski, J.; Papp, R.; Dietrich, E.; Liu, B.; Vallee, F.; Leclaire, M. E.; Fourtounis, J.; Martino, G.; Perryman, A. L.; Pau, V.; Yin, S. Y.; Mader, P.; Roulston, A.; Truchon, J. F.; Marshall, C. G.; Diallo, M.; Duffy, N. M.; Stocco, R.; Godbout, C.; Bonneau-Fortin, A.; Kryczka, R.; Bhaskaran, V.; Mao, D.; Orlicky, S.; Beaulieu, P.; Turcotte, P.; Kurinov, I.; Sicheri, F.; Mamane, Y.; Gallant, M.; Black, W. C., Discovery of an Orally Bioavailable and Selective PKMYT1 Inhibitor, RP-6306. *J Med Chem* **2022,** *65* (15), 10251-10284.

10. Inghardt, T.; Antonsson, T.; Ericsson, C.; Hovdal, D.; Johannesson, P.; Johansson, C.; Jurva, U.; Kajanus, J.; Kull, B.; Michaelsson, E.; Pettersen, A.; Sjogren, T.; Sorensen, H.; Westerlund, K.;

Lindstedt, E. L., Discovery of AZD4831, a Mechanism-Based Irreversible Inhibitor of Myeloperoxidase, As a Potential Treatment for Heart Failure with Preserved Ejection Fraction. *J Med Chem* **2022,** *65* (17), 11485-11496.
11. Chen, B.; Wu, J.; Yan, Z.; Wu, H.; Gao, H.; Liu, Y.; Zhao, J.; Wang, J.; Yang, J.; Zhang, Y.; Pan, J.; Ling, Y.; Wen, H.; Huang, Z., 1,3-Substituted beta-Carboline Derivatives as Potent Chemotherapy for the Treatment of Cystic Echinococcosis. *J Med Chem* **2023,** *66* (24), 16680-16693.
12. Mo, C.; Xu, X.; Zhang, P.; Peng, Y.; Zhao, X.; Chen, S.; Guo, F.; Xiong, Y.; Chu, X. J.; Xu, X., Discovery of HPG1860, a Structurally Novel Nonbile Acid FXR Agonist Currently in Clinical Development for the Treatment of Nonalcoholic Steatohepatitis. *J Med Chem* **2023,** *66* (14), 9363-9375.
13. Hill, M. D.; Blanco, M. J.; Salituro, F. G.; Bai, Z.; Beckley, J. T.; Ackley, M. A.; Dai, J.; Doherty, J. J.; Harrison, B. L.; Hoffmann, E. C.; Kazdoba, T. M.; Lanzetta, D.; Lewis, M.; Quirk, M. C.; Robichaud, A. J., SAGE-718: A First-in-Class N-Methyl-d-Aspartate Receptor Positive Allosteric Modulator for the Potential Treatment of Cognitive Impairment. *J Med Chem* **2022,** *65* (13), 9063-9075.
14. Garnsey, M. R.; Smith, A. C.; Polivkova, J.; Arons, A. L.; Bai, G.; Blakemore, C.; Boehm, M.; Buzon, L. M.; Campion, S. N.; Cerny, M.; Chang, S. C.; Coffman, K.; Farley, K. A.; Fonseca, K. R.; Ford, K. K.; Garren, J.; Kong, J. X.; Koos, M. R. M.; Kung, D. W.; Lian, Y.; Li, M. M.; Li, Q.; Martinez-Alsina, L. A.; O'Connor, R.; Ogilvie, K.; Omoto, K.; Raymer, B.; Reese, M. R.; Ryder, T.; Samp, L.; Stevens, K. A.; Widlicka, D. W.; Yang, Q.; Zhu, K.; Fortin, J. P.; Sammons, M. F., Discovery of the Potent and Selective MC4R Antagonist PF-07258669 for the Potential Treatment of Appetite Loss. *J Med Chem* **2023,** *66* (5), 3195-3211.